

Improved ENSO Forecasting Using Bayesian Updating and the North American Multimodel Ensemble (NMME)

WEI ZHANG AND GABRIELE VILLARINI

IHR–Hydrosience and Engineering, The University of Iowa, Iowa City, Iowa

LOUISE SLATER

IHR–Hydrosience and Engineering, The University of Iowa, Iowa City, Iowa, and Department of Geography, Loughborough University, Loughborough, United Kingdom

GABRIEL A. VECCHI

Department of Geosciences, Princeton University, Princeton, New Jersey

A. ALLEN BRADLEY

IHR–Hydrosience and Engineering, The University of Iowa, Iowa City, Iowa

(Manuscript received 7 February 2017, in final form 25 July 2017)

ABSTRACT

This study assesses the forecast skill of eight North American Multimodel Ensemble (NMME) models in predicting Niño-3/-3.4 indices and improves their skill using Bayesian updating (BU). The forecast skill that is obtained using the ensemble mean of NMME (NMME-EM) shows a strong dependence on lead (initial) month and target month and is quite promising in terms of correlation, root-mean-square error (RMSE), standard deviation ratio (SDRatio), and probabilistic Brier skill score, especially at short lead months. However, the skill decreases in target months from late spring to summer owing to the spring predictability barrier. When BU is applied to eight NMME models (BU-Model), the forecasts tend to outperform NMME-EM in predicting Niño-3/-3.4 in terms of correlation, RMSE, and SDRatio. For Niño-3.4, the BU-Model outperforms NMME-EM forecasts for almost all leads (1–12; particularly for short leads) and target months (from January to December). However, for Niño-3, the BU-Model does not outperform NMME-EM forecasts for leads 7–11 and target months from June to October in terms of correlation and RMSE. Last, the authors test further potential improvements by preselecting “good” models (BU-Model-0.3) and by using principal component analysis to remove the multicollinearity among models, but these additional methodologies do not outperform the BU-Model, which produces the best forecasts of Niño-3/-3.4 for the 2015/16 El Niño event.

1. Introduction

The El Niño–Southern Oscillation (ENSO) phenomenon is a dominant atmospheric–oceanic mode in the tropical Pacific with a dominant time scale of 2–7 years (e.g., Philander 1983; Rasmusson and Wallace 1983; Trenberth 1997; Wyrтки 1975), strongly mediating global

weather and climate (e.g., Alexander et al. 2002; Hoerling et al. 1997; Rasmusson and Wallace 1983; Wang et al. 2000; Webster and Yang 1992; Wyrтки 1973; Zhang et al. 2013). The predictability of the global climate system strongly depends on the prediction of ENSO, which is the largest source of predictability for North Atlantic and Pacific climate, for U.S. precipitation, and for the Asian summer monsoon (e.g., Kumar et al. 2017; Xue et al. 2013; Zhu et al. 2013). It is therefore crucial to advance our understanding and to make timely and reliable forecasts of ENSO.

In recent decades, major advancements have been made in understanding and forecasting ENSO (e.g.,

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-17-0073.s1>.

Corresponding author: Wei Zhang, wei-zhang-3@uiowa.edu

Cane et al. 1986; Battisti and Sarachik 1995; Clarke 2008; L'Heureux and Thompson 2006; Philander 1983; Sarachik and Cane 2010; Stuecker et al. 2015; Wittenberg et al. 2014; Jia et al. 2015), due to the improved capability of fully coupled climate models (e.g., Bellenger et al. 2014; Capotondi 2013; Collins 2000; Delworth et al. 2012; Vecchi and Wittenberg 2010), better atmospheric and oceanic observations (e.g., McPhaden et al. 1998; White 1995; Xie 2004), and improved assimilation techniques to feed observations into climate models (e.g., Behringer et al. 1998; Chen et al. 1995; Jin et al. 2008; Latif et al. 1998; Zhang et al. 2007). However, the predictability of ENSO by climate models is still limited by error growth and model inadequacies (Jin et al. 2008; Kumar et al. 2017; Xue et al. 2013). For example, in early 2014, the forecasts using climate models or statistical methods falsely predicted an El Niño in the 2014/15 winter (Ludescher et al. 2014; Tollefson 2014), and a number of studies have attempted to understand the underlying physical mechanisms for the failure of the 2014/15 case (e.g., Hu and Fedorov 2016; Imada et al. 2016; Min et al. 2015; Zhu et al. 2016).

The North American Multimodel Ensemble (NMME) project (Kirtman et al. 2014) has advanced the forecasting of ENSO and relevant climate variables by integrating coupled models from research centers across the United States and Canada. Kumar et al. (2017) assessed the predictability of Niño-3.4 in the NMME models. They found that the predictability of ENSO strongly depends on seasonality, due to changes in ENSO's predictable component, and is the lowest in spring and summer because of the spring predictability barrier (e.g., Webster and Yang 1992). Although the prediction skill based on the ensemble mean of NMME models is promising, it is of central importance to examine whether we can further improve the NMME forecasts by using more advanced statistical methods to leverage the information from these models. For instance, even though the NMME models have different numbers of ensemble members ranging from 6 to 28, the focus of previous studies was on the use of the ensemble average (weighted equally) to produce the final forecasts (Becker et al. 2014; Chen et al. 2017; Kirtman et al. 2014; Kumar et al. 2017). Whether it is possible to improve the forecast skill by using all the individual members (rather than their ensemble average) has not been examined in previous studies.

From a methodological perspective, Bayesian updating (BU) has proven skillful in improving multimodel forecasts and provides a more realistic description of predictive uncertainty accounting for between- and in-model variances (Bradley et al. 2015; Duan et al. 2007; Hoeting et al. 1999; Luo and Wood 2008; Min et al. 2007;

Raftery et al. 2005; Slater et al. 2017). BU implements Bayes's theorem to update the probability distribution of a variable (e.g., NMME-based Niño-3.4 forecasts) with the new observed information (e.g., observation-based Niño-3.4). The BU predictions are basically weighted averages of the individual forecasts of climate variables (Luo and Wood 2008; Luo et al. 2007; Wang et al. 2013). BU has been used to improve ENSO forecasts with the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble simulations (Coelho et al. 2004). We will use BU to further improve the NMME forecasting of ENSO by leveraging the forecasting skill of all of the individual members from eight NMME models.

The objectives of this study are twofold. First, we aim to evaluate the skill of the NMME models in predicting Niño-3/-3.4 indices. Second, we attempt to further improve the NMME forecasts for Niño-3/-3.4 indices by leveraging the forecasting skill of eight NMME models using BU. We evaluate the prediction of the 2015/16 El Niño event using BU and compare the forecasts with the NMME models. This study aims to advance our understanding of the current status of the skill of the NMME models in forecasting ENSO and provides a new approach to improve forecasts using ensemble members, which has potential to be broadly applied.

The remainder of this paper is organized as follows. Section 2 presents the data and methodology, while section 3 discusses the forecast results based on NMME and BU. Section 4 includes the discussion and summarizes the main conclusions.

2. Data and methodology

a. NMME models

The available period, ensemble size, and lead months of the NMME models are summarized in Table 1, with eight climate models and up to 94 members (Becker et al. 2014; Kirtman et al. 2014). The hindcasts and forecasts of sea surface temperature (SST) at $1^\circ \times 1^\circ$ spatial resolution are available from the early 1980s to the present. We consider eight climate models: Community Climate System Model, version 3 (CCSM3), and Community Climate System Model, version 4, subset of CESM1 (CCSM4), from the National Center for Atmospheric Research (NCAR), Center for Ocean-Land-Atmosphere Studies (COLA), and Rosenstiel School of Marine and Atmospheric Science, University of Miami (RSMAS); Third Generation Canadian Coupled Global Climate Model (CanCM3) and Fourth Generation Canadian Coupled Global Climate Model (CanCM4) from Environment Canada's Meteorological Service of Canada-Canadian Meteorological Centre

TABLE 1. The available period, ensemble size, and lead months of the NMME climate models. The available period does not reflect the presence of gaps; the ensemble size indicates the largest number of members per model and does not reflect missing data for one or more members.

Model	Available period	Ensemble size	Lead times (months)
CCCma CanCM3	1981–present	10	0.5–11.5 (1–12)
CCCma CanCM4	1981–present	10	0.5–11.5 (1–12)
COLA–RSMAS CCSM3	1982–present	6	0.5–11.5 (1–12)
COLA–RSMAS CCSM4	1982–present	10	0.5–11.5 (1–12)
NCEP CFSv2	1982–present	24	0.5–9.5 (1–10)
GFDL CM2.1	1982–present	10	0.5–11.5 (1–12)
GFDL FLOR B01	1980–present	12	0.5–11.5 (1–12)
NASA GEOS5	1981–present	12	0.5–9.5 (1–10)
NCAR CESM1	1980–2010 and 2016–present	10	0.5–11.5 (1–12)

(CMC); operational Climate Forecast System, version 2, (CFSv2) from the National Centers for Environmental Prediction (NCEP); Goddard Earth Observing System Model, version 5 (GEOS5), from the National Aeronautics and Space Administration (NASA)'s Global Modeling and Assimilation Office (GMAO); Geophysical Fluid Dynamics Laboratory Climate Model, version 2.1 (GFDL CM2.1), and Forecast-Oriented Low Ocean Resolution version of CM2.5 (FLOR B01) from National Oceanic and Atmospheric Administration (NOAA)/GFDL. The observed estimates of SST are obtained from the Met Office Hadley Centre (HadISST, version 1.1) (Rayner et al. 2003).

b. Niño-3 and Niño-3.4 indices

We focus on the Niño-3 and Niño-3.4 indices in both the observations and NMME hindcasts/forecasts. These two indices are defined as the SST anomalies averaged over the Niño-3 and Niño-3.4 regions. The Niño-3 region is bounded by 5°S–5°N and 150°–90°W, while the Niño-3.4 region is bounded by 5°S–5°N and 170°–120°W. The SST anomalies in the observations are calculated by removing the seasonal cycle, which is based on the climatology of 1982–2015. The SST anomalies in the NMME models are calculated by accounting for the dependence on season and on forecast lead time with respect to the 1982–2015 period following Kumar et al. (2017).

c. BU

The BU of the NMME forecasts is an implementation of Bayes's theorem, in which the probability distribution of a variable Y (i.e., NMME-based Niño-3.4 forecasts) is updated when new information (e.g., observation-based Niño-3.4) becomes available. The BU-Model is defined as the method that directly applies BU to the eight models in NMME. BU has also been used to improve the ensemble forecasts of the ECMWF model (Coelho et al. 2004), where it was applied to calibrate and combine both empirical and coupled model ensemble

forecasts. This study employs BU to combine coupled model ensemble forecasts from the NMME project. The best estimates of the probability of different outcomes are defined by the climatology (i.e., the historical averages of the forecasted variable), represented here by the prior climatological density function $f(y)$. After a climate model forecast θ is available, the updated (or posterior) density function is given by Bayes's theorem to be

$$f(y|\theta) = \frac{f_{\theta}(\theta|y)f(y)}{f_{\theta}(\theta)}, \quad (1)$$

where $f_{\theta}(\theta)$ is the unconditional density of θ , and $f_{\theta}(\theta|y)$ is the likelihood function. The posterior density $f(y|\theta)$ describes the conditional distribution of the variable given the climate model forecast θ and therefore represents a probability distribution forecast of the outcome. Here we apply Bayesian updating to a data sample, where y_i ($i = 1, \dots, N$) represents the historical observations of Y [i.e., a sample drawn from the prior density $f(y)$]. We represent a sample drawn from the posterior density $f(y|\theta)$ using the likelihood function $f_{\theta}(\theta|y)$. By definition, the likelihood function $f_{\theta}(\theta|y)$ is the distribution of a given model forecast θ (e.g., July 2010) conditioned on the observed SST y for the same month. If we have a hindcasted sample (e.g., monthly observations from January 1982 to December 2015), the likelihood function can be estimated by a regression model:

$$\theta = \bar{\theta}(y) + \varepsilon, \quad (2)$$

where $\bar{\theta}(y)$ is the expected value of the forecast given the observation y , and ε is the residual model error. We apply the Bayesian updating using a linear regression approach, so we implement a simple linear regression model $\bar{\theta}(y)$ and assume that the residual errors are normally distributed with constant variance (see also Coelho et al. 2004). The likelihood function $f_{\theta}(\theta|y)$ is then

$$f_{\theta}(\theta | y) = \frac{1}{\sqrt{2\pi}\sigma_{\varepsilon}} \exp\left\{-\frac{[\bar{\theta}(y) - \theta]^2}{2\sigma_{\varepsilon}^2}\right\}. \quad (3)$$

Using the likelihood function developed for each of the 94 individual model members or the ensemble average of the eight models, we assign a weight w_i to each observation y_i in the historical sample. The weight w_i represents the likelihood of observing outcome y_i given the climate forecast θ . The historical sample is re-weighted as follows:

$$w_i = \frac{f_{\theta}(\theta | y_i)}{\sum_{j=1}^N f_{\theta}(\theta | y_j)}, \quad (4)$$

where the sum of the weights w_i is equal to 1. The collection of the weights for all historical observations for the given month (e.g., for the HadISST-derived Niño-3.4 from January 1870 to December 2015, minus the forecast year) is thus similar to a discrete probability distribution forecast for each model or model member. This suggests that the weights show the likelihood of each discrete outcome given the climate model forecasts. Weights of $1/N$ indicate that there is no potential skill and produce the same distribution as the prior distribution before Bayesian updating, so the output is equivalent to a climatology forecast (i.e., the average historical conditions for the same months) and the member is automatically ignored. For models with a weak relationship between forecasts and observations, the Bayesian weights will be close to $1/N$, indicating that each outcome is equally likely. For models with a strong and significant relationship between forecasts and observations, the Bayesian weights will be greater than $1/N$ and will grow as the potential skill increases (thus giving more weight to the forecast). Any weights of less than $1/N$ indicate that the outcome is less likely than the climatology. The weights for every single model are combined to yield a multimodel forecast.

The BU method may be dependent on the skill of the individual models of the NMME project. Here we assess whether the skill of the BU can be improved by selecting the models (“good candidates”) in which the value of the correlation coefficient between the forecasted and observed Niño-3/-3.4 indices is greater than a threshold value, selected to be 0.3 in this study (BU-Model-0.3). The threshold of 0.3 is selected experimentally based on the value of correlation coefficient at the 0.05 significance level for the study period. We tuned the threshold to be larger or smaller, and the results did not change significantly. Moreover, because we cannot assume that all the GCMs are independent (i.e., similarities exist among different models, and some, like

TABLE 2. Description of the four ways in which BU is applied in this study.

BU methods	Description
BU-Model	BU is applied to the eight models in NMME.
BU-Model-0.3	BU is applied to the NMME models for which the correlation with observed El Niño indices is greater than 0.3.
BU-Model-PCA	BU is applied to the loadings of PCs for the eight models in NMME.
BU-Model-PCA-0.3	BU is applied to the loadings of PCs for the eight models for which the correlation with observed El Niño indices is greater than 0.3.

CCSM3/CCSM4 or CanCM3/CanCM4, are two different versions of the same model), we perform principal component analysis (PCA) on the forecasted Niño-3/-3.4 indices for each NMME model to reduce the multicollinearity among individual forecasts [see also [Slater et al. \(2017\)](#) for an application to the seasonal forecasting of precipitation and temperature over Europe using the NMME data]. Because the forecasts of the Niño indices with different NMME models are linearly correlated, the PCA, which transforms the forecasts into orthogonal principal components, may improve the forecasts by reducing the multicollinearity. We then apply BU to the loadings of all the PCs (BU-Model-PCA). Similar to BU-Model-0.3, we also focus the Bayesian updating on the loadings of the PCs having correlation with the observed Niño-3/-3.4 indices greater than 0.3 (BU-Model-PCA-0.3). The four BU methods are summarized in [Table 2](#).

d. Forecast verification metrics

To quantify the skill of the different models and approaches with respect to the observations we use the correlation coefficient, the root-mean-square error (RMSE), and the standard deviation ratio (SDratio) as deterministic metrics. The SDratio represents the capability of the forecasts in capturing the dispersion of the observations ([Barnston et al. 2015](#)) and is defined as the standard deviation of the forecasted El Niño indices divided by the one for the observations.

We will refer to the correlation between the mean forecasted Niño indices (of all NMME members) and observations (NMME-CorM) and to the mean of all the individual correlations between every NMME-member and the observations (NMME-MCor) (i.e., the correlation of the means vs the mean of the correlations). To be consistent with the calculation of other skills (e.g., RMSE and SDratio), we also refer to the method based

on the ensemble mean of all NMME members/models (NMME-EM). NMME-CorM is a special case of NMME-EM for calculating correlation. Although NMME-CorM and NMME-MCor are forecast verification measures, we use them as forecast methods to be easily compared with BU hereafter.

In addition to deterministic verification metrics such as correlation and RMSE, we also employed the probabilistic verification metric Brier skill score (BSS) (Wilks 2011). The Brier skill score is based on the Brier score (BS), which is a scalar metric of the accuracy of a probabilistic forecast for dichotomous events and is defined as follows:

$$BS = \frac{1}{n} \sum_{i=1}^n (f_i - O_i)^2, \quad (5)$$

where n is the number of forecasts, f_i is the forecast probability of the occurrence of an event for the i th forecast, and O_i is the i th observed probability, which is defined to be 1 if the event occurs and 0 if it does not.

The Brier skill score is defined as follows:

$$BSS = 1 - \frac{BS_f}{BS_{ci}}, \quad (6)$$

where BS_{ci} denotes the Brier score for climatological forecast (with a probability of 0.33 for each tercile), while BS_f is the Brier score for the forecast based on NMME or BU. For a climatological forecast, the BSS is zero. In this study, a probabilistic forecast of an event in each tercile was implemented. The three categories are defined as “above normal,” “normal,” and “below normal” based on the values of the Niño-3/-3.4 index in the forecasts. We focus on the forecast skill of above normal and below normal events with warm/cold SST anomalies.

To test whether the differences in forecast skill among the different forecast methods are statistically significant, we use the Wilcoxon signed-rank test. This test considers the magnitude of the differences in forecast skills (DelSole and Tippett 2014), and its statistic is defined as follows:

$$WT = \sum_{n \text{ when } d_n > 0} (\text{rank of } |d_n|), \quad (7)$$

where $d_n = 0$ is assumed to never occur. The finite-sample distribution of this statistic is invariant to the distribution of the loss differential if the distribution is symmetric about zero.

3. Results

Figure 1 shows the observed and predicted composite SST anomalies for the December–February (DJF) El

Niño and La Niña events between 1981 and 2016 based on observations and NMME forecasts initialized in December. Overall, the NMME climate models successfully predicted the SST anomalies in the tropical Pacific, especially in the Niño-3 and Niño-3.4 regions. During El Niño years, the forecast SST anomalies in the Niño regions are slightly weaker than the observations, with the exception of the GFDL CM2.1 and the NASA GEOS5 climate models. During La Niña years, the SST anomalies in the NMME models tend to extend farther west than the observations, with the exception of CFSv2, which predicts the SST anomalies at locations similar to the observations. The negative SST anomalies in GFDL CM2.1, CanCM3, CanCM4, CCSM3, and NASA GEOS5 models are slightly stronger than those in the observations (Fig. 1).

The model biases during El Niño and La Niña years are shown in Fig. S1 (see the online supplemental material), supporting the above discussions on SST anomalies. For example, GFDL CM2.1 has warm biases in the tropical Pacific, especially west of the Niño-3.4 region during El Niño years. CanCM3, CanCM4, and CCSM3 also show weak warm biases west of the Niño-3.4 region. Most of the models show cold biases in the Niño-3 and Niño-3.4 regions during El Niño years. During La Niña years, most of the models feature cold biases in the Niño-3 and Niño-3.4 regions except CCSM4, with warm biases in the Niño-3.4 region.

Figure 2 displays the temporal evolution of the Niño-3.4 index during El Niño and La Niña years. Overall, the NMME models capture the temporal evolution of the El Niño/La Niña periods quite well, though there are some biases in all the NMME models. Given such biases in the SST anomalies and evolution, the BU could overcome some of these biases by integrating useful information from historical observations.

NMME-CorM shows higher correlation value for Niño-3/-3.4 than NMME-MCor (Figs. 3 and 4), indicating that the ensemble mean of NMME forecasts is better than individual NMME forecasts. Figure 3 shows the skill of the Niño-3 forecasts for target months from January to December and lead months from 1 to 12 with different forecast methods. Overall, BU performs better in target months from January to March than in other target months, consistent with the changes of forecast skill of NMME-CorM with respect to different target months. Overall, NMME-CorM performs better in target months during boreal autumn and winter than in target months during spring and summer in terms of correlation (Fig. 3), consistent with Barnston et al. (2015) and Kumar et al. (2017), who found that the predictability of ENSO is the lowest in spring and summer. The skill of NMME Niño-3 forecasts depends

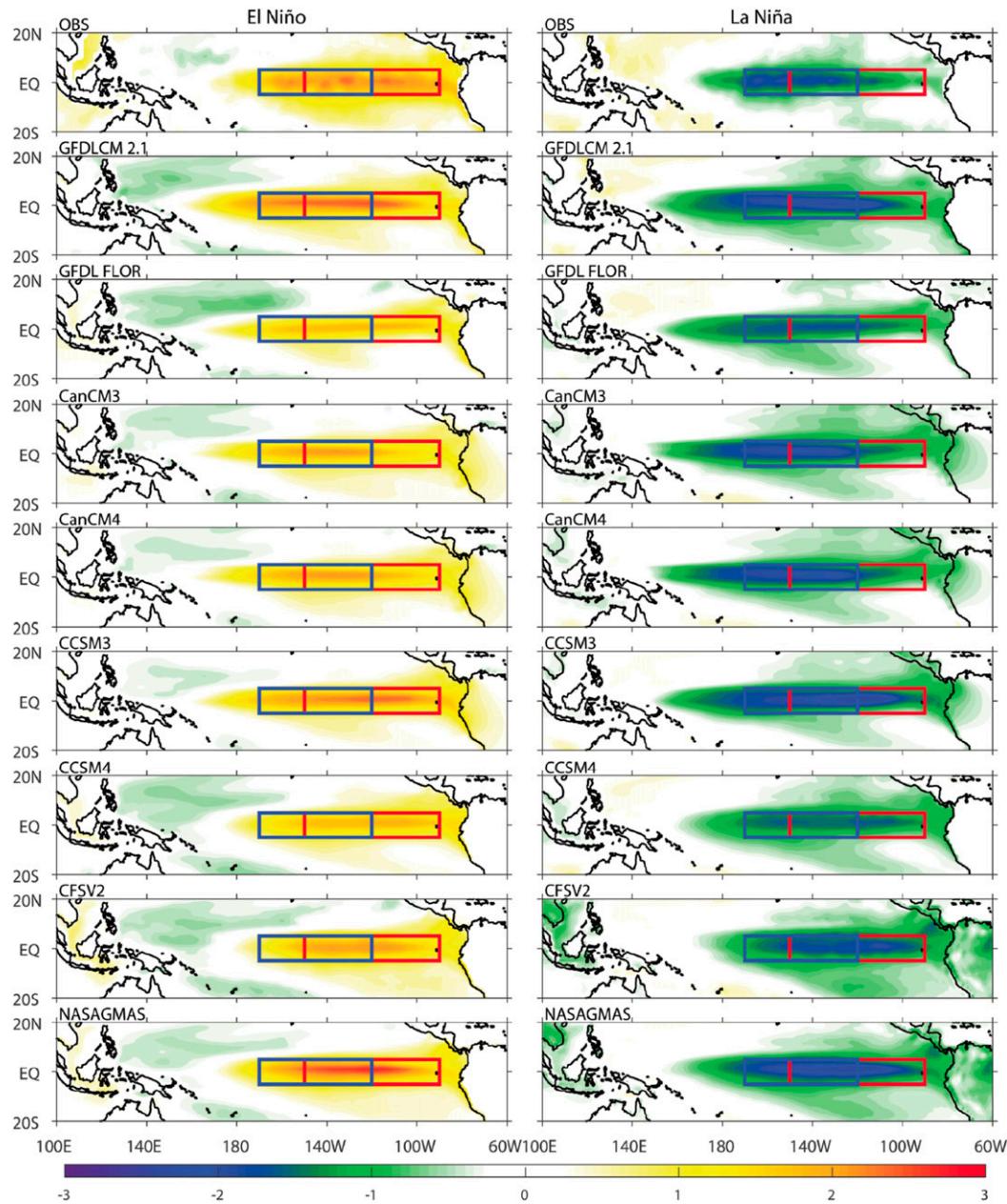


FIG. 1. Composite SST anomalies (shading; $^{\circ}\text{C}$) for the DJF (left) El Niño events (1982/83, 1986/87, 1991/92, 1997/98, 2002/03, 2009/10, 2015/16) and (right) La Niña events (1984/85, 1988/89, 1995/96, 1998/99, 1999/2000, 2007/08, 2010/11) in (top) observations and NMME forecasts initialized in December. The blue and red rectangles represent Niño-3.4 and Niño-3 regions, respectively.

on the lead month, and the skill of NMME-CorM drops with increasing lead time (Fig. 3). For example, the forecast skill of NMME (i.e., NMME-CorM) for Niño-3 in terms of correlation is close to 1 (e.g., 0.95–0.98) at lead month 1 but drops to ~ 0.6 for target months from January to April and to ~ 0.4 for target months from May to December at lead month 12 (still significant at the 5% significance level). This reduction in forecast

skill with increasing lead time suggests that NMME-CorM produces promising results for the Niño-3 index in terms of correlation, with a strong dependence on lead month and target month (Barnston and Tippett 2013; Barnston et al. 2012, 2015; Jin et al. 2008; Kumar et al. 2017; Tippett et al. 2012).

Figure 3 also shows the forecasts of Niño-3 in all target and initialization months using the four BU methods

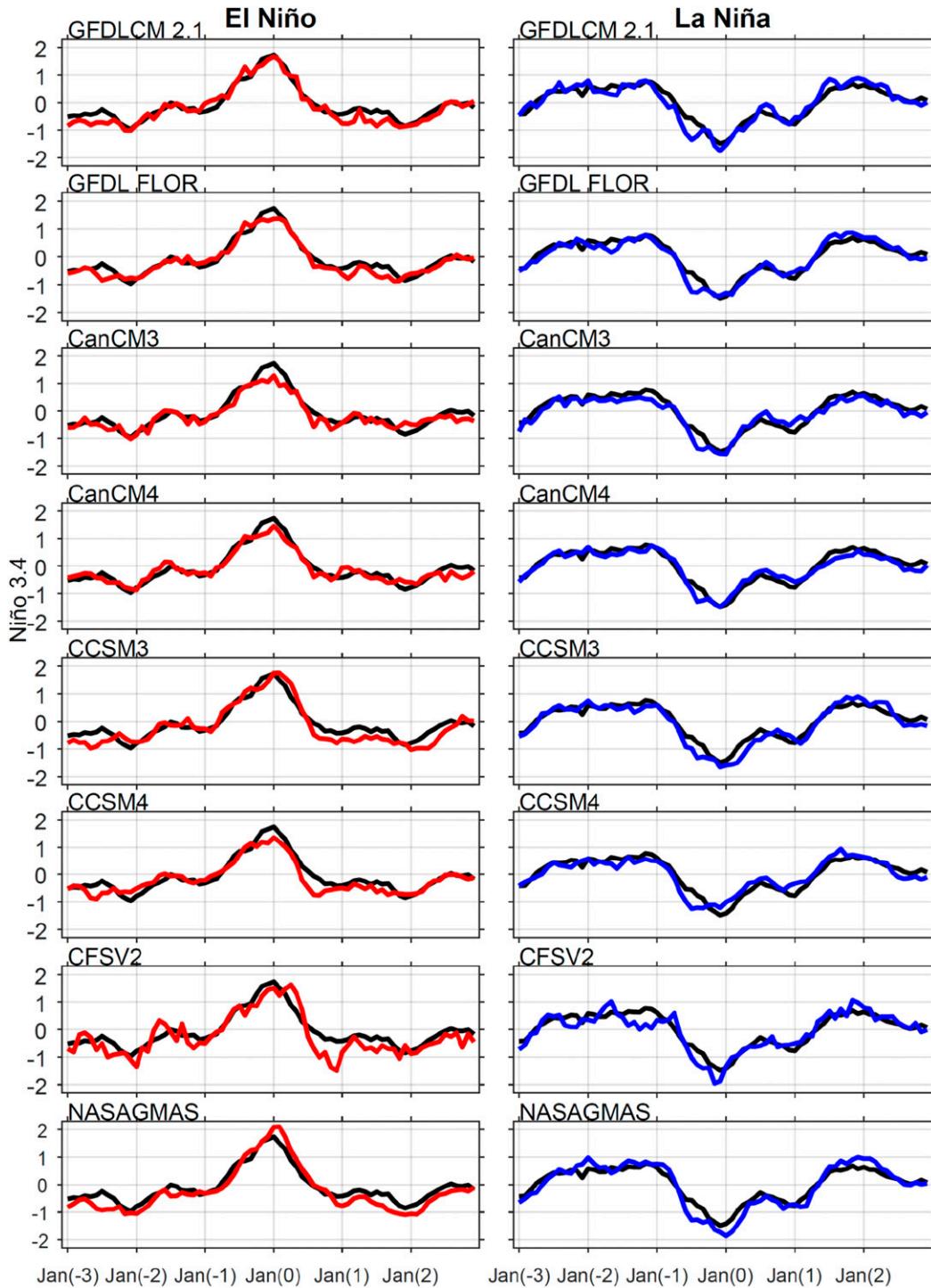


FIG. 2. Evolution of Niño-3.4 in observations (black) and in the NMME forecasts with climate models for El Niño (red) and La Niña (blue) events. Jan(0) represents the January in mature El Niño/La Niña phases.

listed in Table 2. For target months from January to May, the BU-Model generally outperforms the skill of NMME-CorM/NMME-EM. As expected, the forecast skill in both NMME and BU forecasts drops for

increasing lead times. Similar to the skill obtained from the equal weighting of the NMME forecasts, the forecast skill with the BU-Model drops at a slower rate in target months from January to April than in target months

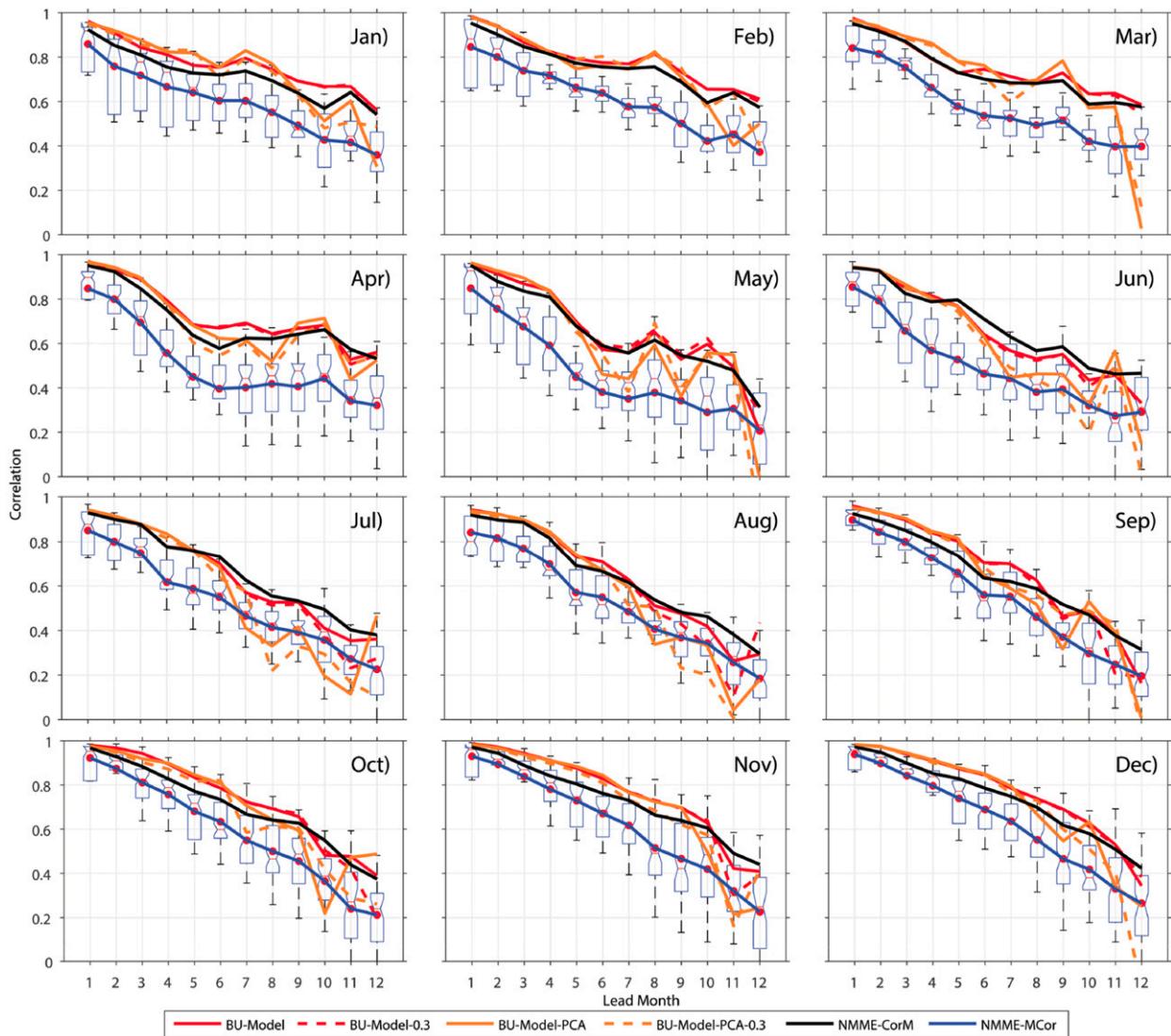


FIG. 3. Skill of the Niño-3 forecasts for target months from January to December and lead months from 1 (0.5) to 12 (11.5) with BU-Model (red solid), BU-Model-0.3 (red dashed), BU-Model-PCA (orange solid), and BU-Model-PCA-0.3 (orange dashed). NMME-CorM (black solid line) denotes the skill of the ensemble mean (equally weighted) of all the NMME members, while NMME-MCor (blue solid) denotes the mean skill of all NMME members. The y axis indicates the strength of the correlation between the forecasts and the observed Niño indices. In the box plots, the circle and the red line within the box represent the mean and the median, respectively; the limits of the box represent the 25th and 75th percentiles, while the limits of the whiskers are the 5th and 95th percentiles.

from May to December when the lead month increases from 1 to 12. For example, the forecast skill of the BU-Model for Niño-3 in terms of correlation is around 0.95–0.98 at lead month 1, around 0.6 for target month from January to April, and around 0.4 at lead month 12 for May–December target months. For June–July target months, the BU-Model performs slightly better than NMME-CorM for short lead months 1–4, after which the opposite is true. BU-Model-PCA is developed by applying BU to the loadings of all the PCs and aims to improve the forecast skill of BU by removing

collinearity (see section 2 for details). In these target months, the BU-Model-PCA does not outperform the BU-Model. For August–December target months, the BU-Model generally performs better than the NMME-CorM in forecasting Niño-3 (Fig. 3).

We build the BU-Model based on the NMME forecasts with a correlation of 0.3 or higher (BU-Model-0.3) to test whether we can improve the performance of BU by keeping only the forecasts having a higher correlation with observations. BU-Model-0.3 does not perform better than the BU-Model for almost any of the target

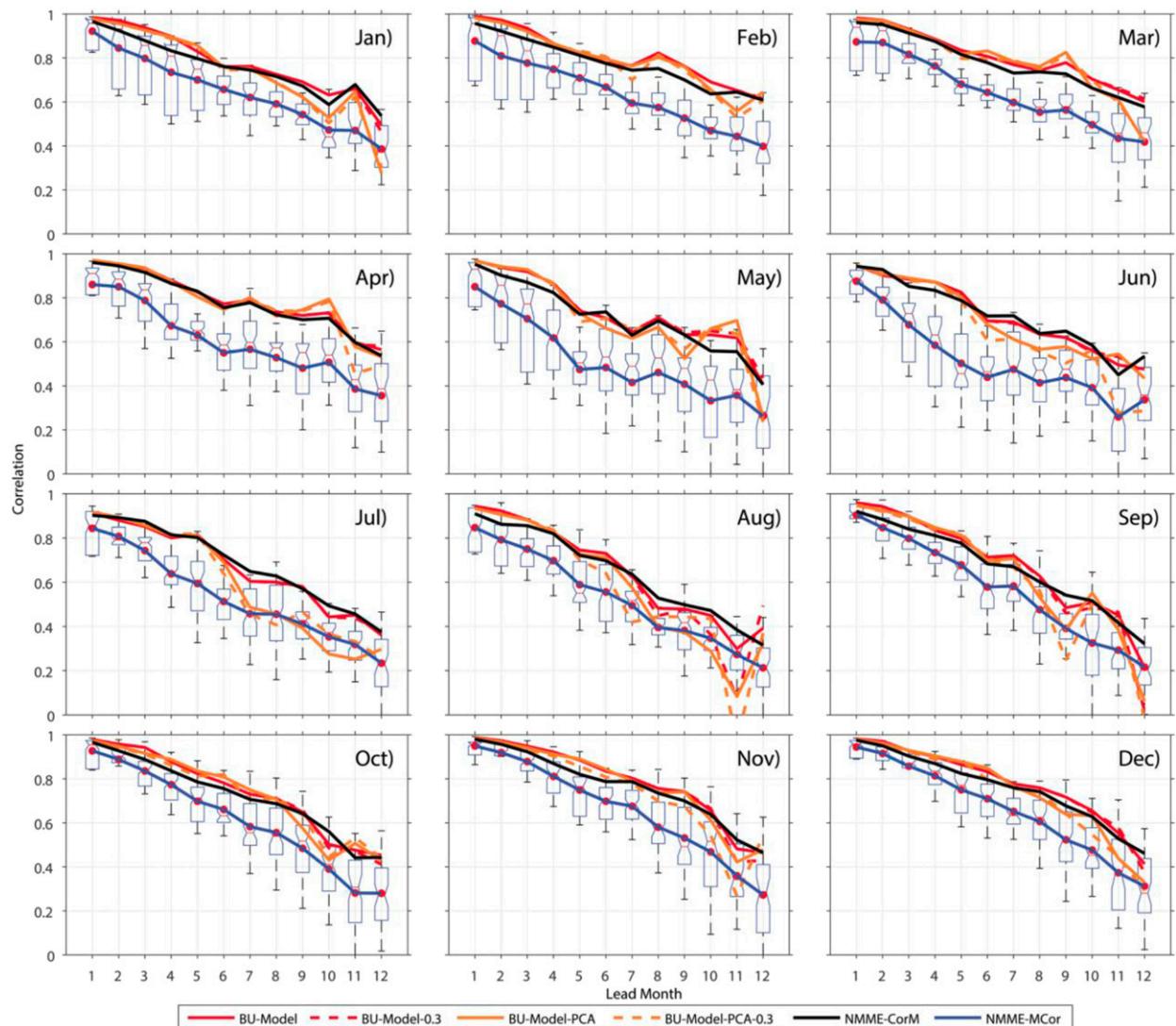


FIG. 4. As in Fig. 3, but for the Niño-3.4 index.

months (Fig. 3), and this statement holds for different threshold values of the correlation coefficient (figure not shown). Thus, the skill does not improve by focusing on a subset of models that exhibit a stronger relationship between forecasts and observations. The BU-Model-PCA/BU-PCA-Model-0.3 can slightly outperform BU-Model/BU-Model-0.3 for very short lead months (Fig. 3). However, BU-Model/BU-Model-0.3 performs much better than the BU-Model-PCA/BU-PCA-Model-0.3 for longer lead months (Fig. 3). This suggests that the application of PCA to the NMME forecasts prior to BU does not lead to a consistent improvement in the forecast skill for all lead months.

Figure 4 shows the results for Niño-3.4, which are similar to those shown for the Niño-3 index (Fig. 3). For

example, the BU-Model and BU-Model-0.3 outperform the NMME-CorM in target months January–May and September–December (Fig. 4). However, for June–August target months, the BU-Model/BU-Model-0.3 performs better than NMME-CorM only for short lead months. Previous studies have shown the difficulties in forecasting ENSO during June–August because of the spring predictability barrier, which presents a challenge (Barnston and Tippett 2013; Barnston et al. 2012, 2015; McPhaden 2003; Tippett et al. 2012). The spring predictability barrier is responsible for the drop in forecast skill during boreal spring and for the drop in skill of the forecasts made during boreal spring for the following seasons. For example, the skill of forecasts initialized in boreal spring drops faster than that of the forecasts

initialized in August or November (Jin et al. 2008). Previous studies have reported low forecast skill for target months June–August initialized during spring (e.g., Torrence and Webster 1998; Jin et al. 2008; Barnston et al. 2015). Because the BU-Model-0.3 and BU-Model-PCA-0.3 do not outperform the BU-Model and BU-Model-PCA, respectively, we will focus on the performance of BU-Model and BU-Model-PCA in forecasting Niño-3/-3.4 henceforth.

The forecast skill (correlation) for Niño-3/-3.4 using the BU-Model generally outperforms NMME-CorM for January–May and September–December target months (Figs. 3 and 4). For June–August target months, the BU-Model does not show improvements in forecast skill for Niño-3/-3.4 with respect to NMME-CorM. Figure 5 shows a summary of the values of the correlation coefficient between forecasted and observed Niño indices using NMME-CorM, BU-Model, and BU-Model-PCA. The spring predictability barrier in forecasting El Niño/La Niña is evident for all the forecasts, with diminished skill for target months from late boreal spring to boreal summer. The dependence of skill on season and lead month is also obvious for all the forecasts. Overall, the BU-Model outperforms NMME-CorM in forecasting Niño-3/-3.4, particularly for short lead months 1–5. In general, the BU-Model-PCA does not outperform BU-Model in forecasting Niño-3/-3.4 (Fig. 5). Figure S2 illustrates the differences in forecast skill between BU-Model/BU-Model-PCA and NMME-CorM for Niño-3.4 and Niño-3. In general, the BU-Model outperforms NMME-CorM for almost all short lead months. However, the BU-Model shows some weakness in forecasting Niño-3.4/-3 for June–August target months after lead month 5. BU-Model-PCA shows similar results compared with the BU-Model for short lead months. However, the BU-Model-PCA performs worse than the BU-Model for August–September target months and lead months 10–12 for Niño-3.4 and for June–November target months and lead months 10–12 for Niño-3. The largest improvements in forecasting Niño-3/-3.4 made by BU-Model/BU-Model-PCA compared with NMME-CorM lie in the August–December target months and lead months 1–7.

The SDratio measures the ability of the forecasts to capture the dispersion of the observations. We find the NMME-EM forecasts tend to be overdispersed in comparison with the observed values for long lead months, especially in the January–June and October–December target months (Fig. 6, top). Moreover, the NMME-EM forecasts tend to underestimate the dispersion in the observations for lead months 1–4 and July–October target months. The BU-Model largely outperforms the NMME-EM forecasts in terms of

SDratio for almost all lead and target months because the SDratio values in BU-Model are closer to 1 (Fig. 6, middle). BU-Model-PCA improves the SDratio by reducing dispersion for very short lead months but increases dispersion for the longest lead months 10–12 (Fig. 6, bottom). This suggests that BU-Model-PCA does not improve the forecasting of ENSO compared with the BU-Model.

The RMSE values of the NMME-EM forecasts of Niño-3/-3.4 tend to decrease as the lead month becomes shorter (Fig. 7, top). The largest RMSE occurs in the lead months 11–12 and the target months October–December (OND). For Niño-3.4, there are large RMSE values in the NMME-EM forecasts in the January target month with 10–12 lead months (Fig. 8, top). Overall, the RMSE is smaller for the BU-Model than for NMME-EM, particularly for the short lead months, consistent with the improvements in correlation in Figs. 5 and 7 (middle). Moreover, the RMSE in BU-Model-PCA is also slightly smaller than that in NMME-EM for short lead months (Fig. 7, bottom). To support the above discussions, Fig. S3 illustrates the differences in the RMSE values between the BU-Model and NMME-EM and between BU-Model-PCA and NMME-EM. Overall, BU outperforms NMME-EM by producing a smaller RMSE in Niño-3/-3.4 indices, especially for short lead months; BU-Model-PCA produces similar RMSE compared with BU-Model. It is noted that BU-Model performs much better than BU-Model-PCA in terms of correlation coefficient between forecasted and observed Niño indices for long lead months (Figs. 4, 5, and S2). However, the differences in RMSE between BU-Model and BU-Model-PCA appear to be smaller compared with the difference in correlation.

We use the Wilcoxon signed-rank test (DelSole and Tippett 2014) to test whether the differences in the forecast skill (e.g., correlation and RMSE) between BU-Model and NMME are statistically significant. Figure 8 shows that the differences in RMSE for Niño-3.4 between BU-Model and NMME are statistically significant at the 5% level for lead months 1–5, identical to those between BU-Model-PCA and NMME. For the forecasts of Niño-3, the differences in RMSE between BU and NMME are significantly significant for lead months 1–6, and this is also true for differences between BU-Model-PCA and NMME. For the correlation coefficient, the differences in Niño-3.4 between BU-Model and NMME are statistically significant at lead months 1–5, and this is also true for the differences between BU-Model-PCA and NMME (Niño-3.4), between BU-Model and NMME for Niño-3, and between BU-Model-PCA and NMME for Niño-3. There are some significant differences between BU-Model-PCA and NMME for Niño-3 and Niño-3.4 at lead month 12.

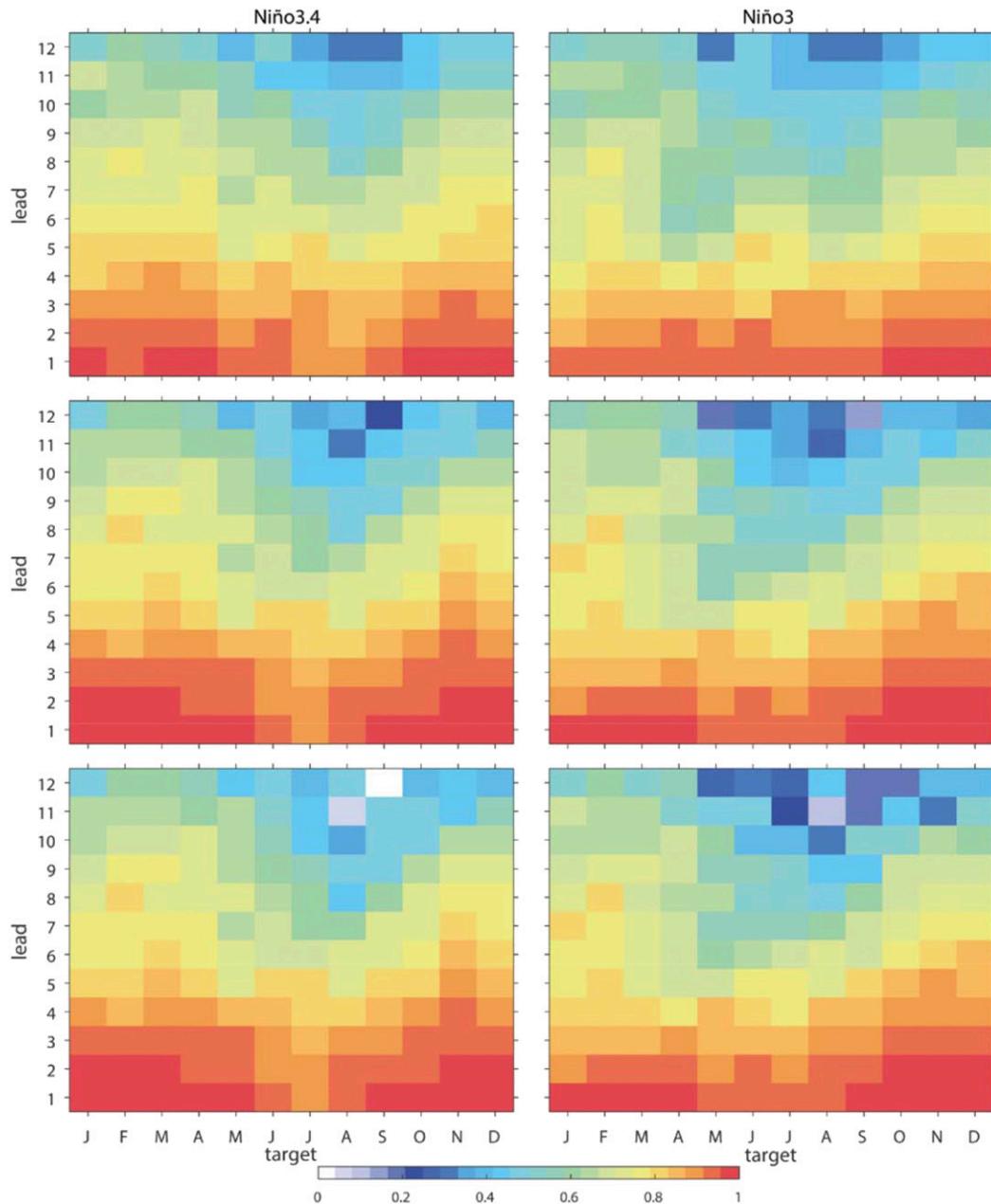


FIG. 5. The correlation values (shading) in (top) NMME-CorM/NMME-EM, (middle) BU-Model, and (bottom) BU-Model-PCA for (left) Niño-3.4 and (right) Niño-3.

In addition to the deterministic measures of skill (i.e., correlation and RMSE), we also use Brier skill score to measure the forecast skill with BU-Model and NMME. We focus on the forecasts for the upper and lower terciles of the Niño-3.4 and Niño-3 indices. Overall, the BU method outperforms NMME for the upper tercile of Niño-3.4 and Niño-3 at shorter lead months (Fig. 9) and for the lower tercile, especially at short lead months (Fig. 10). Overall, the forecast skill for the lower tercile of Niño-3.4 is higher than for Niño-3 (Fig. 10).

The 2015/16 El Niño event is one of the strongest El Niño events since 1870 (Blunden and Arndt 2016). Here we use this El Niño event as a case study to show the capability of the BU-Model in forecasting the Niño-3/-3.4 indices. We focus on the observed Niño-3/-3.4 index averaged over OND in 2015 with a value of 2.6. The Niño-3/-3.4 indices forecasted by the BU-Model are generally much closer to the observations than those forecasted by NMME-EM up to the lead month 10 (Fig. 11). The forecasts of Niño-3/-3.4 indices

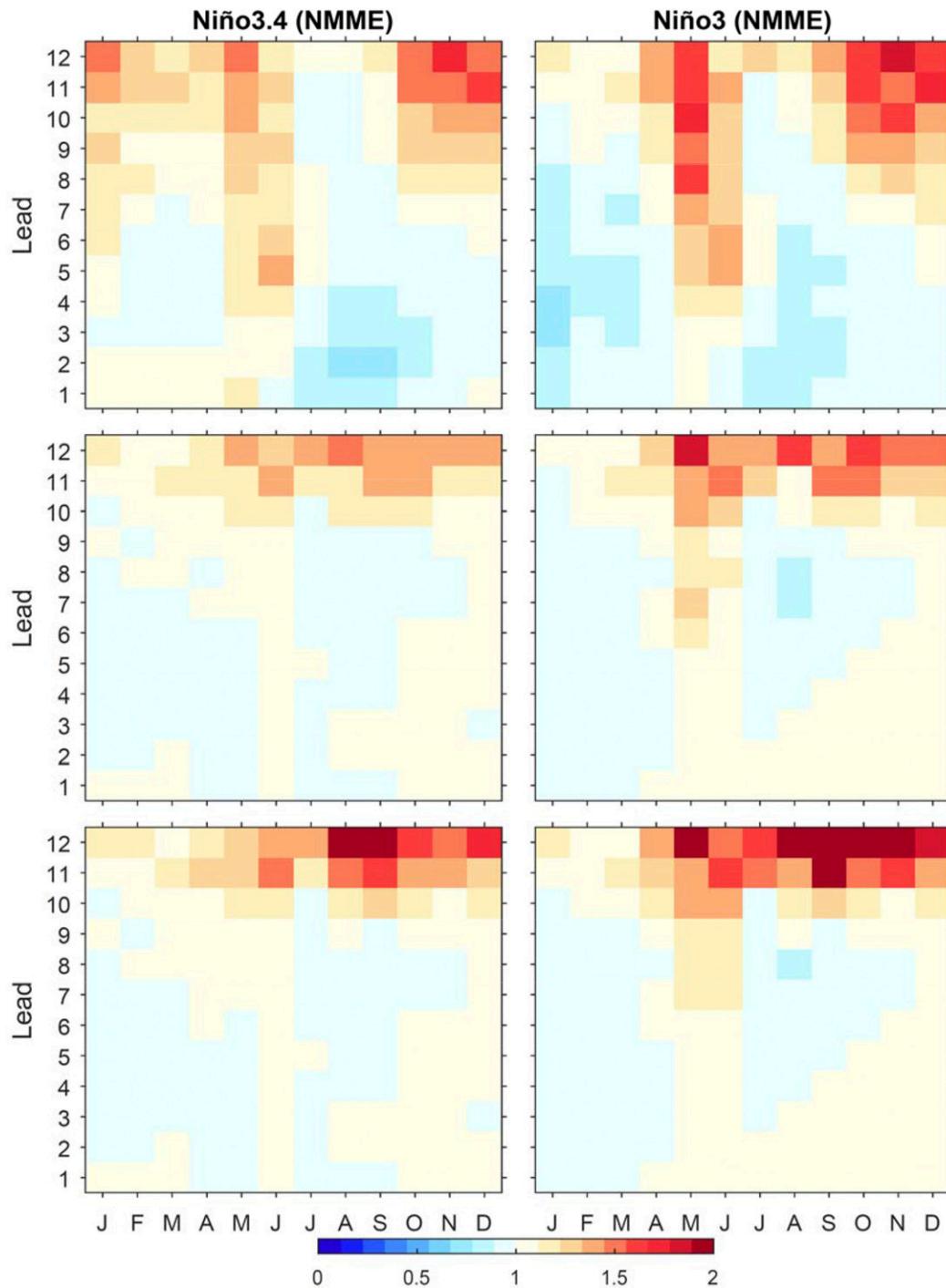


FIG. 6. SDRatio (shading) using forecasts/observations in (top) NMME-EM, (middle) BU-Model, and (bottom) BU-Model-PCA for (left) Niño-3.4 and (right) Niño-3. A value of 1 is the ideal value.

obtained during the OND of 2015/16 with the BU-Model also produce much smaller biases than those achieved by NMME-EM even for lead month 10. In the lead month 9 the forecasted Niño-3 index with NMME-EM is ~ 0.5 while the index obtained with the BU-Model is ~ 1.3 . In

the lead month 8, the forecasted Niño-3 index during OND 2015 with NMME-EM is ~ 1 while the forecasted Niño 3 with BU-Model is ~ 2 . Therefore, the BU-Model performs much better than NMME-EM in forecasting this most recent strong El Niño event.

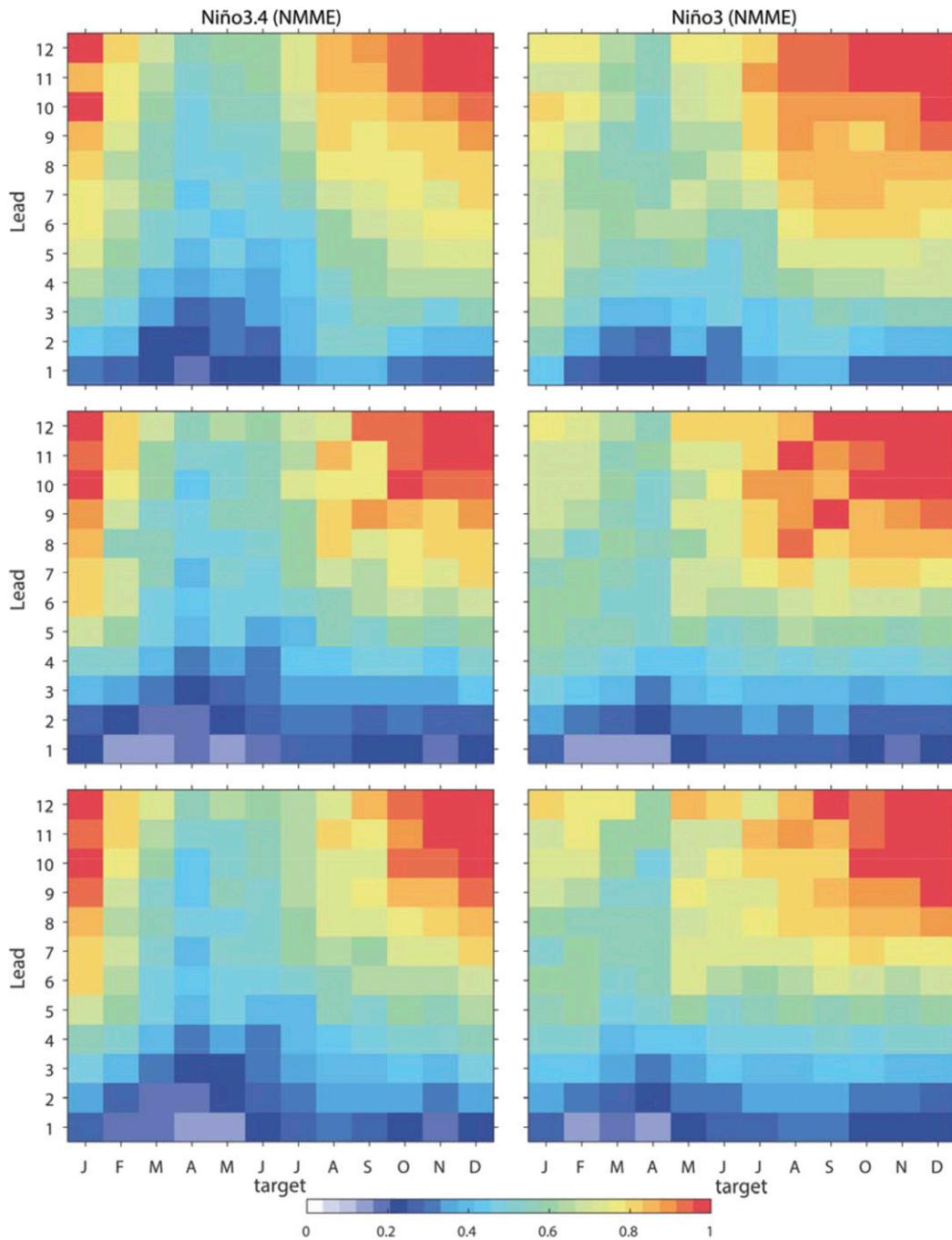


FIG. 7. RMSE (shading) in (top) NMME-EM, (middle) BU-Model, and (bottom) BU-Model-PCA for (left) Niño-3.4 and (right) Niño-3.

We have also examined whether longer or more reliable observations of SST can influence the forecast skill of the El Niño events (e.g., 1982/83, 1997/98, and 2015/16). To accomplish this, we use the observed SST over the period 1870–2015 (BU-1870) and 1940–2015 (BU-1940) for the BU for the three El Niño events (Fig. S4). There are some differences in the forecasts

based on BU-1870 and BU-1940; overall, BU-1870 produces better forecast skill than BU-1940 for the 2015/16 El Niño event (Fig. S4), while the skill of BU-1940 and BU-1870 for the 1997/98 and 1982/83 events appears to be similar. Based on these analyses, we cannot find an obvious improvement in the prediction skill obtained from BU-1870 compared with that from

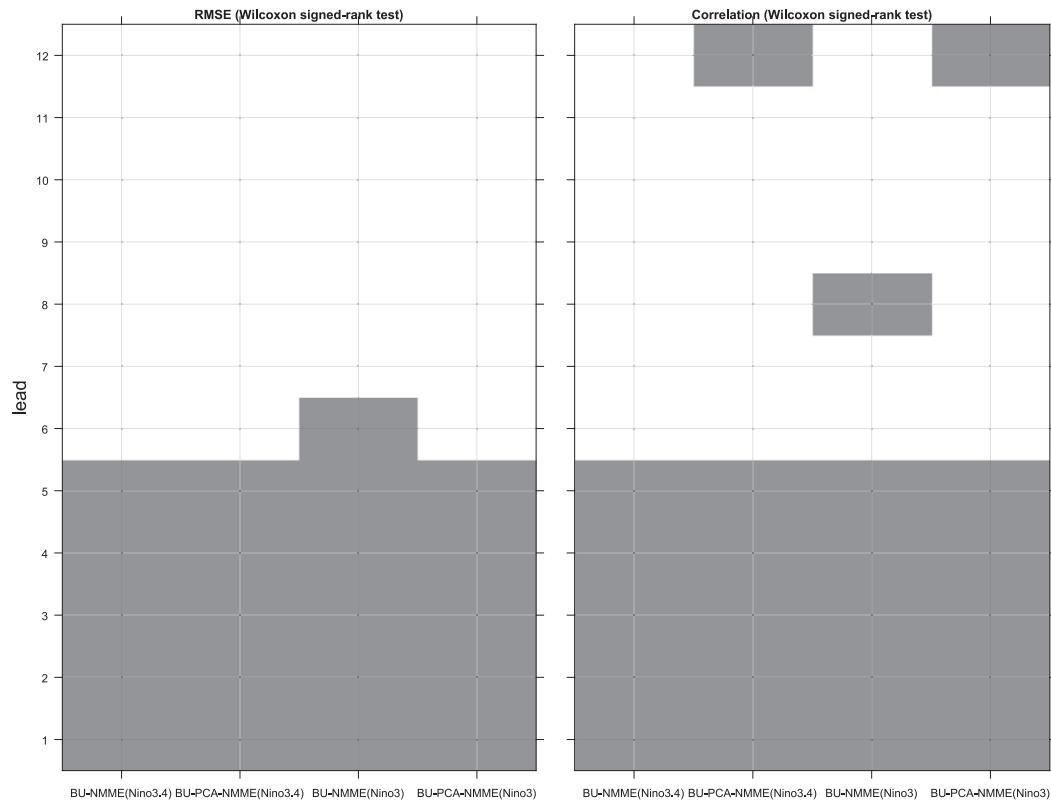


FIG. 8. Wilcoxon signed-rank test results for (left) RMSE and (right) correlation coefficient in different BU versions and NMME for Niño-3 and Niño-3.4 indices at lead months 1–12. Dark-shaded regions indicate that the differences are statistically significant.

BU-1940. Future studies should examine this issue in more detail.

4. Discussion and conclusions

Timely and accurate ENSO forecasts are likely to have major societal and economic impacts (e.g., agriculture and fishing). The NMME project has advanced our capability of forecasting key atmospheric and oceanic variables. In this study, we have assessed the ability of a Bayesian updating (BU) approach to improve the forecasts of the Niño-3/-3.4 indices and compared the results with those of the equally weighted ensemble average of the NMME forecasts (NMME-EM).

The forecast skill for Niño-3/-3.4 using NMME-EM shows strong dependence on lead (initial) month and target month and is promising in terms of correlation, RMSE, and SDRatio, especially at short lead months. For example, the correlation coefficient between forecasted (NMME-EM) and observed Niño-3/-3.4 indices is close to 1 for short lead months, with very small RMSE errors. The skill in terms of correlation for Niño-3/-3.4 with NMME-EM drops to 0.4–0.7 at lead month 12. Moreover, the forecast skill of Niño-3/-3.4 with

NMME-EM decreases in target months from late spring to summer, due to the spring predictability barrier.

Overall, the BU-Model outperforms NMME-EM in predicting Niño-3/-3.4 for almost all the target months and lead months in terms of correlation, RMSE, and SDRatio. The BU-Model outperforms NMME-EM forecasts in Niño-3.4 for almost all lead months and target months. However, it does show some weaknesses in forecasting Niño-3.4/-3 for June–August target months and long lead months (e.g., 7–10) in terms of correlation, and it does not outperform the NMME forecasts for Niño-3 for June–October target months and lead months 7–11. A caveat of this study is that the ENSO forecast can vary over decadal scales (Barnston et al. 2012; Zhao et al. 2016), and we would need to study longer periods to obtain more robust comparisons between Bayesian updating and NMME. For the 2014/15 event, the prediction skill of Niño-3.4 during October–December with BU-Model is higher than NMME-EM for lead months 1–5 and 10, consistent with a better skill of BU at shorter lead months (Fig. S5). Overall, the BU-Model performs better than NMME-EM for Niño-3/-3.4 in terms of SDRatio, but focusing on a subset of GCMs that showed a stronger relationship between forecasts

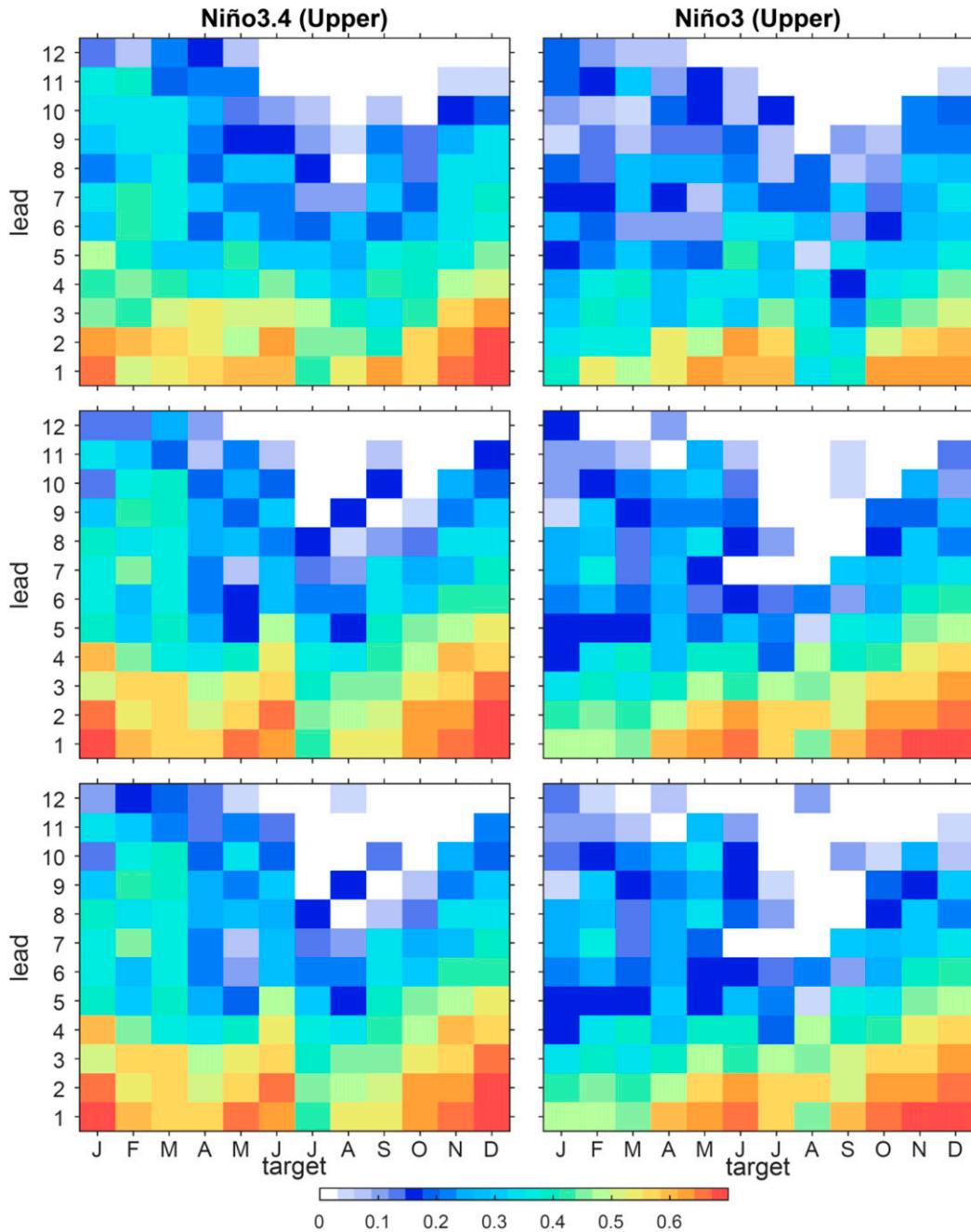


FIG. 9. BSS for the upper tercile in (top) NMME-EM, (middle) BU-Model, and (bottom) BU-Model-PCA for (left) Niño-3.4 and (right) Niño-3.

and observations does not improve the overall forecast skill. We also used PCA to examine the potential impacts of correlation among models on forecast skill, but we did not find any significant improvement. Meanwhile, the performance of BU in forecasting Niño-4 (Fig. S6) is slightly worse than for Niño-3/-3.4. This might be due to the intrinsic low predictability of central Pacific El Niño events. Further studies are required to

improve the BU for Niño-4, possibly by using a more sophisticated likelihood function and prior distribution.

The prediction skill that is obtained using BU-Model is comparable to, or slightly better than, that of current prediction models/schemes for Niño-3/-3.4. For example, at a 3-month lead time in the eastern equatorial Pacific, an intermediate coupled climate model produces a correlation of ~ 0.75 (Zheng and Zhu 2016),

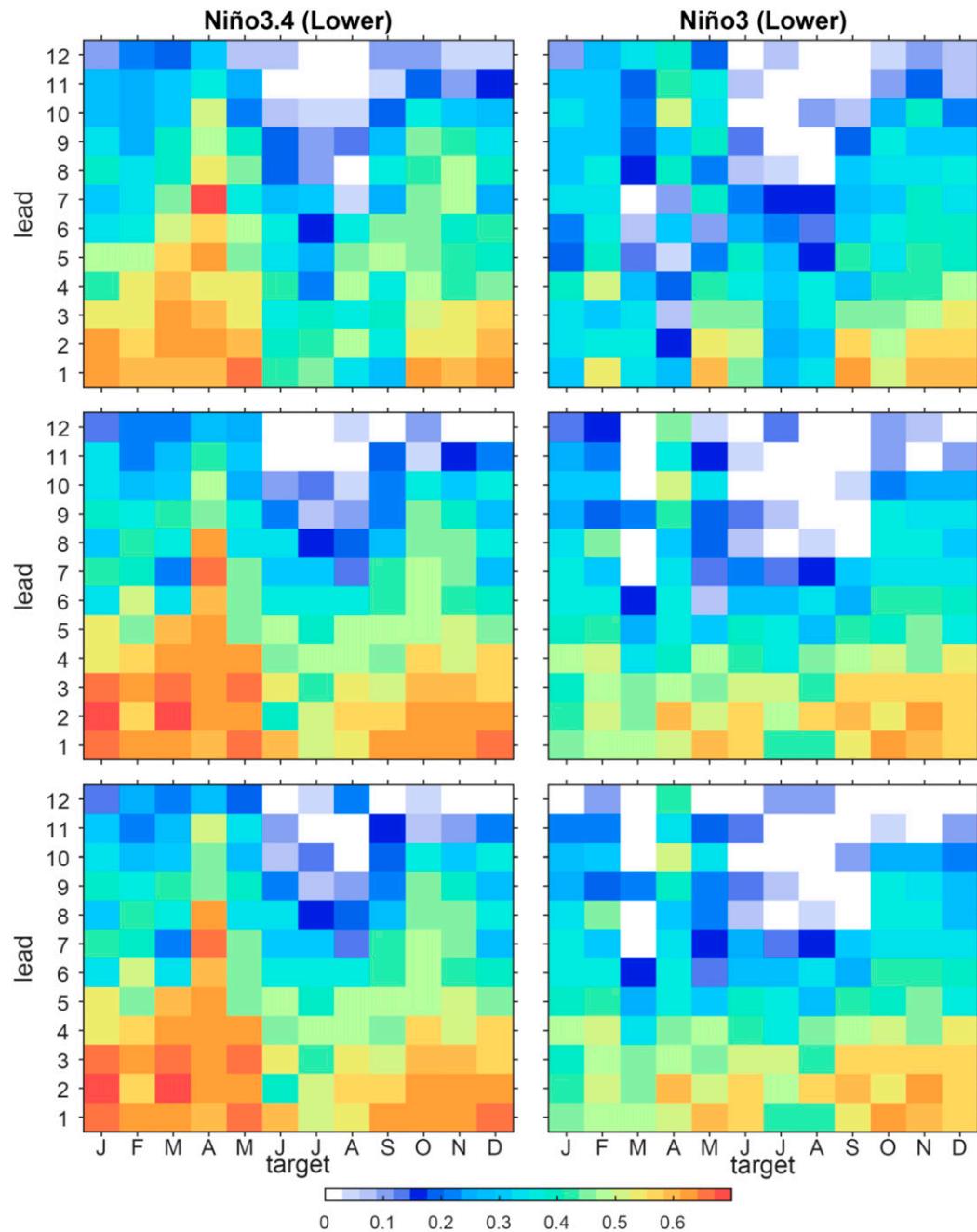


FIG. 10. As in Fig. 9, but for the lower tercile.

while the NMME-EM and BU-Model produce a correlation of more than 0.85. The BU-Model forecast skill is comparable to the skill that is obtained using CFSv1 and CFSv2 (with biases adjusted) as reported in Barnston and Tippett (2013). Our forecasts of Niño-3.4 are also comparable to or slightly better than those obtained in Saha et al. (2014), particularly for short lead months. Our BU-Model slightly outperforms the results based on statistical models and dynamic models

reported in Barnston et al. (2012) in terms of correlation and RMSE.

ENSO plays a central role in exciting teleconnections that modulate global weather and climate, such as tropical cyclones, precipitation, and air temperature (e.g., Alexander et al. 2002; Ropelewski and Halpert 1986; Zhang et al. 2012, 2015, 2016a,b). Thus, our future work will assess whether the improvements in the prediction skill of Niño-3/-3.4 indices using BU-Model can

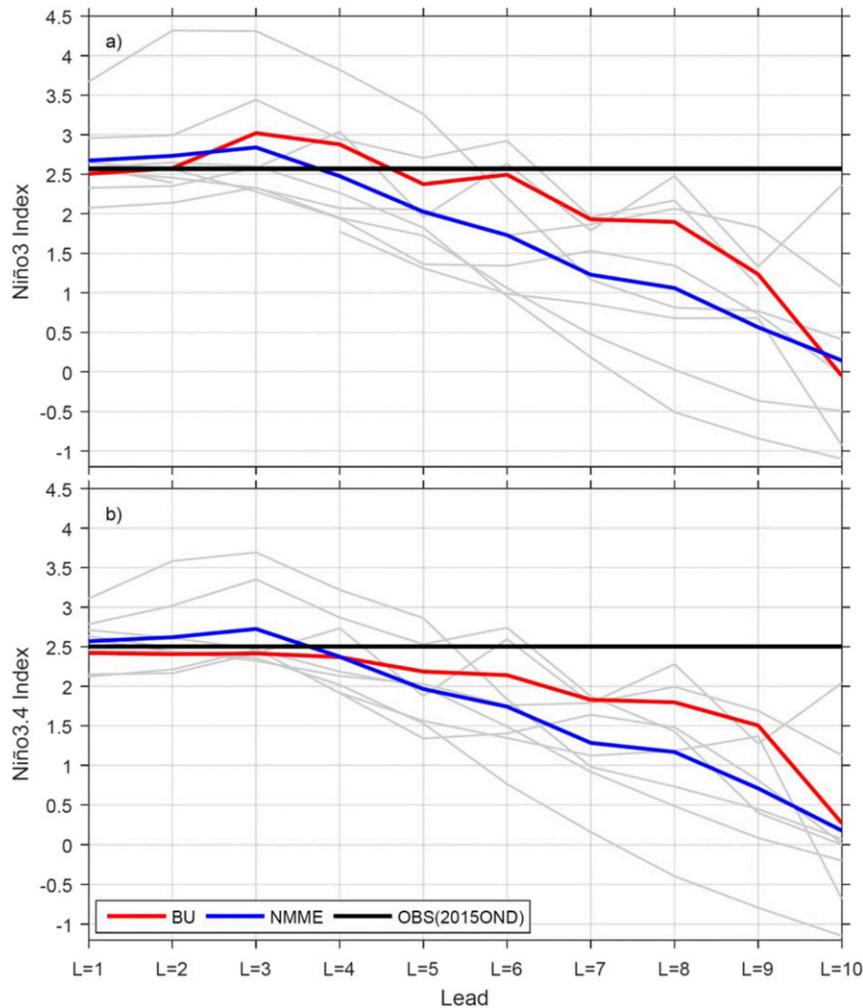


FIG. 11. The observed values (black) and forecasts with BU (red) and NMME (blue) for (a) Niño-3 and (b) Niño-3.4 index averaged over OND of the 2015/16 El Niño. The gray curves represent the Niño-3/-3.4 indices that were predicted by each of the eight available NMME models (mean of the members).

heighten the prediction skill of these ENSO-driven meteorological variables and phenomena.

Acknowledgments. The authors thank the NMME program partners and acknowledge the help of NCEP, IRI, and NCAR personnel in creating, updating, and maintaining the NMME archive, with the support of NOAA, NSF, NASA, and DOE. This study was partly supported by NOAA's Climate Program Office's Modeling, Analysis, Predictions, and Projections Program, Grant NA15OAR4310073, and Award NA14OAR4830101 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce. GV also acknowledges funding by the National Science Foundation under CAREER Grant AGS-1349827 and the Broad Agency Announcement

Program and the Engineer Research and Development Center–Cold Regions Research and Engineering Laboratory under Contract W913E5-16-C-0002. The authors acknowledge insightful comments by three anonymous reviewers, which improved the quality of this paper.

REFERENCES

- Alexander, M. A., I. Bladé, M. Newman, J. R. Lanzante, N.-C. Lau, and J. D. Scott, 2002: The atmospheric bridge: The influence of ENSO teleconnections on air–sea interaction over the global oceans. *J. Climate*, **15**, 2205–2231, doi:[10.1175/1520-0442\(2002\)015<2205:TABTIO>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<2205:TABTIO>2.0.CO;2).
- Barnston, A. G., and M. K. Tippett, 2013: Predictions of Niño3.4 SST in CFSv1 and CFSv2: A diagnostic comparison. *Climate Dyn.*, **41**, 1615–1633, doi:[10.1007/s00382-013-1845-2](https://doi.org/10.1007/s00382-013-1845-2).

- , —, M. L. L'Heureux, S. Li, and D. G. DeWitt, 2012: Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bull. Amer. Meteor. Soc.*, **93**, 631–651, doi:[10.1175/BAMS-D-11-00111.1](https://doi.org/10.1175/BAMS-D-11-00111.1).
- , —, H. M. van den Dool, and D. A. Unger, 2015: Toward an improved multimodel ENSO prediction. *J. Appl. Meteor. Climatol.*, **54**, 1579–1595, doi:[10.1175/JAMC-D-14-0188.1](https://doi.org/10.1175/JAMC-D-14-0188.1).
- Battisti, D., and E. Sarachik, 1995: Understanding and predicting ENSO. *Rev. Geophys.*, **33**, 1367–1376, doi:[10.1029/95RG00933](https://doi.org/10.1029/95RG00933).
- Becker, E., H. van den Dool, and Q. Zhang, 2014: Predictability and forecast skill in NMME. *J. Climate*, **27**, 5891–5906, doi:[10.1175/JCLI-D-13-00597.1](https://doi.org/10.1175/JCLI-D-13-00597.1).
- Behringer, D. W., M. Ji, and A. Leetmaa, 1998: An improved coupled model for ENSO prediction and implications for ocean initialization. Part I: The ocean data assimilation system. *Mon. Wea. Rev.*, **126**, 1013–1021, doi:[10.1175/1520-0493\(1998\)126<1013:AICMFE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1013:AICMFE>2.0.CO;2).
- Bellenger, H., É. Guilyardi, J. Leloup, M. Lengaigne, and J. Vialard, 2014: ENSO representation in climate models: From CMIP3 to CMIP5. *Climate Dyn.*, **42**, 1999–2018, doi:[10.1007/s00382-013-1783-z](https://doi.org/10.1007/s00382-013-1783-z).
- Blunden, J., and D. S. Arndt, Eds., 2016: State of the Climate in 2015. *Bull. Amer. Meteor. Soc.*, **97** (8), S1–S275, doi:[10.1175/2016BAMSStateoftheClimate.1](https://doi.org/10.1175/2016BAMSStateoftheClimate.1).
- Bradley, A. A., M. Habib, and S. S. Schwartz, 2015: Climate index weighting of ensemble streamflow forecasts using a simple Bayesian approach. *Water Resour. Res.*, **51**, 7382–7400, doi:[10.1002/2014WR016811](https://doi.org/10.1002/2014WR016811).
- Cane, M. A., S. E. Zebiak, and S. C. Dolan, 1986: Experimental forecasts of El Niño. *Nature*, **321**, 827–832, doi:[10.1038/321827a0](https://doi.org/10.1038/321827a0).
- Capotondi, A., 2013: ENSO diversity in the NCAR CCSM4 climate model. *J. Geophys. Res. Oceans*, **118**, 4755–4770, doi:[10.1002/jgrc.20335](https://doi.org/10.1002/jgrc.20335).
- Chen, D., S. E. Zebiak, A. J. Busalacchi, and M. A. Cane, 1995: An improved procedure for El Niño forecasting: Implications for predictability. *Science*, **269**, 1699–1702, doi:[10.1126/science.269.5231.1699](https://doi.org/10.1126/science.269.5231.1699).
- Chen, L.-C., H. van den Dool, E. Becker, and Q. Zhang, 2017: ENSO precipitation and temperature forecasts in the North American Multimodel Ensemble: Composite analysis and validation. *J. Climate*, **30**, 1103–1125, doi:[10.1175/JCLI-D-15-0903.1](https://doi.org/10.1175/JCLI-D-15-0903.1).
- Clarke, A. J., 2008: *An Introduction to the Dynamics of El Niño and the Southern Oscillation*. Academic Press, 324 pp.
- Coelho, C. A. S., S. Pezzulli, M. Balmaseda, F. J. Doblas-Reyes, and D. B. Stephenson, 2004: Forecast calibration and combination: A simple Bayesian approach for ENSO. *J. Climate*, **17**, 1504–1516, doi:[10.1175/1520-0442\(2004\)017<1504:FCACAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<1504:FCACAS>2.0.CO;2).
- Collins, M., 2000: The El Niño–Southern Oscillation in the second Hadley Centre coupled model and its response to greenhouse warming. *J. Climate*, **13**, 1299–1312, doi:[10.1175/1520-0442\(2000\)013<1299:TENOSO>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1299:TENOSO>2.0.CO;2).
- DelSole, T., and M. K. Tippett, 2014: Comparing forecast skill. *Mon. Wea. Rev.*, **142**, 4658–4678, doi:[10.1175/MWR-D-14-00045.1](https://doi.org/10.1175/MWR-D-14-00045.1).
- Delworth, T. L., and Coauthors, 2012: Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *J. Climate*, **25**, 2755–2781, doi:[10.1175/JCLI-D-11-00316.1](https://doi.org/10.1175/JCLI-D-11-00316.1).
- Duan, Q., N. K. Ajami, X. Gao, and S. Sorooshian, 2007: Multimodel ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour.*, **30**, 1371–1386, doi:[10.1016/j.advwatres.2006.11.014](https://doi.org/10.1016/j.advwatres.2006.11.014).
- Hoerling, M. P., A. Kumar, and M. Zhong, 1997: El Niño, La Niña, and the nonlinearity of their teleconnections. *J. Climate*, **10**, 1769–1786, doi:[10.1175/1520-0442\(1997\)010<1769:ENOLNA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<1769:ENOLNA>2.0.CO;2).
- Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky, 1999: Bayesian model averaging: A tutorial. *Stat. Sci.*, **14**, 382–401.
- Hu, S., and A. V. Fedorov, 2016: Exceptionally strong easterly wind burst stalling El Niño of 2014. *Proc. Natl. Acad. Sci. USA*, **113**, 2005–2010, doi:[10.1073/pnas.1514182113](https://doi.org/10.1073/pnas.1514182113).
- Imada, Y., H. Tatebe, M. Watanabe, M. Ishii, and M. Kimoto, 2016: South Pacific influence on the termination of El Niño in 2014. *Sci. Rep.*, **6**, 30341, doi:[10.1038/srep30341](https://doi.org/10.1038/srep30341).
- Jia, L., and Coauthors, 2015: Improved seasonal prediction of temperature and precipitation over land in a high-resolution GFDL climate model. *J. Climate*, **28**, 2044–2062, doi:[10.1175/JCLI-D-14-00112.1](https://doi.org/10.1175/JCLI-D-14-00112.1).
- Jin, E. K., and Coauthors, 2008: Current status of ENSO prediction skill in coupled ocean–atmosphere models. *Climate Dyn.*, **31**, 647–664, doi:[10.1007/s00382-008-0397-3](https://doi.org/10.1007/s00382-008-0397-3).
- Kirtman, B. P., and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, doi:[10.1175/BAMS-D-12-00050.1](https://doi.org/10.1175/BAMS-D-12-00050.1).
- Kumar, A., Z.-Z. Hu, B. Jha, and P. Peng, 2017: Estimating ENSO predictability based on multi-model hindcasts. *Climate Dyn.*, **48**, 39–51, doi:[10.1007/s00382-016-3060-4](https://doi.org/10.1007/s00382-016-3060-4).
- Latif, M., and Coauthors, 1998: A review of the predictability and prediction of ENSO. *J. Geophys. Res.*, **103**, 14 375–14 393, doi:[10.1029/97JC03413](https://doi.org/10.1029/97JC03413).
- L'Heureux, M. L., and D. W. Thompson, 2006: Observed relationships between the El Niño–Southern Oscillation and the extratropical zonal-mean circulation. *J. Climate*, **19**, 276–287, doi:[10.1175/JCLI3617.1](https://doi.org/10.1175/JCLI3617.1).
- Ludescher, J., A. Gozolchiani, M. I. Bogachev, A. Bunde, S. Havlin, and H. J. Schellnhuber, 2014: Very early warning of next El Niño. *Proc. Natl. Acad. Sci. USA*, **111**, 2064–2066, doi:[10.1073/pnas.1323058111](https://doi.org/10.1073/pnas.1323058111).
- Luo, L., and E. F. Wood, 2008: Use of Bayesian merging techniques in a multimodel seasonal hydrologic ensemble prediction system for the eastern United States. *J. Hydrometeorol.*, **9**, 866–884, doi:[10.1175/2008JHM980.1](https://doi.org/10.1175/2008JHM980.1).
- , —, and M. Pan, 2007: Bayesian merging of multiple climate model forecasts for seasonal hydrological predictions. *J. Geophys. Res.*, **112**, D10102, doi:[10.1029/2006JD007655](https://doi.org/10.1029/2006JD007655).
- McPhaden, M. J., 2003: Tropical Pacific Ocean heat content variations and ENSO persistence barriers. *Geophys. Res. Lett.*, **30**, 1480, doi:[10.1029/2003GL016872](https://doi.org/10.1029/2003GL016872).
- , and Coauthors, 1998: The Tropical Ocean-Global Atmosphere observing system: A decade of progress. *J. Geophys. Res.*, **103**, 14 169–14 240, doi:[10.1029/97JC02906](https://doi.org/10.1029/97JC02906).
- Min, Q., J. Su, R. Zhang, and X. Rong, 2015: What hindered the El Niño pattern in 2014? *Geophys. Res. Lett.*, **42**, 6762–6770, doi:[10.1002/2015GL064899](https://doi.org/10.1002/2015GL064899).
- Min, S.-K., D. Simonis, and A. Hense, 2007: Probabilistic climate change predictions applying Bayesian model averaging. *Philos. Trans. Roy. Soc. London*, **365A**, 2103–2116, doi:[10.1098/rsta.2007.2070](https://doi.org/10.1098/rsta.2007.2070).
- Philander, S. G., 1983: El Niño Southern Oscillation phenomena. *Nature*, **302**, 295–301, doi:[10.1038/302295a0](https://doi.org/10.1038/302295a0).
- Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast

- ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174, doi:10.1175/MWR2906.1.
- Rasmusson, E. M., and J. M. Wallace, 1983: Meteorological aspects of the El Niño/Southern Oscillation. *Science*, **222**, 1195–1202, doi:10.1126/science.222.4629.1195.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, doi:10.1029/2002JD002670.
- Ropelewski, C. F., and M. S. Halpert, 1986: North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114**, 2352–2362, doi:10.1175/1520-0493(1986)114<2352:NAPATP>2.0.CO;2.
- Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, doi:10.1175/JCLI-D-12-00823.1.
- Sarachik, E. S., and M. A. Cane, 2010: *The El Niño-Southern Oscillation Phenomenon*. Cambridge University Press, 384 pp.
- Slater, L. J., G. Villarini, and A. A. Bradley, 2017: Weighting of NMME temperature and precipitation forecasts across Europe. *J. Hydrol.*, **552**, 646–659, doi:10.1016/j.jhydrol.2017.07.029.
- Stuecker, M. F., F.-F. Jin, and A. Timmermann, 2015: El Niño–Southern Oscillation frequency cascade. *Proc. Natl. Acad. Sci. USA*, **112**, 13 490–13 495, doi:10.1073/pnas.1508622112.
- Tippett, M. K., A. G. Barnston, and S. Li, 2012: Performance of recent multimodel ENSO forecasts. *J. Appl. Meteor. Climatol.*, **51**, 637–654, doi:10.1175/JAMC-D-11-093.1.
- Tollefson, J., 2014: El Niño tests forecasters. *Nature*, **508**, 20–21, doi:10.1038/508020a.
- Torrence, C., and P. J. Webster, 1998: The annual cycle of persistence in the El Niño/Southern Oscillation. *Quart. J. Roy. Meteor. Soc.*, **124**, 1985–2004, doi:10.1002/qj.49712455010.
- Trenberth, K. E., 1997: The definition of El Niño. *Bull. Amer. Meteor. Soc.*, **78**, 2771–2777, doi:10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2.
- Vecchi, G. A., and A. T. Wittenberg, 2010: El Niño and our future climate: Where do we stand? *Wiley Interdiscip. Rev.: Climate Change*, **1**, 260–270, doi:10.1002/wcc.33.
- Wang, B., R. Wu, and X. Fu, 2000: Pacific–East Asian teleconnection: How does ENSO affect East Asian climate? *J. Climate*, **13**, 1517–1536, doi:10.1175/1520-0442(2000)013<1517:PEATHD>2.0.CO;2.
- Wang, H., B. Reich, and Y. H. Lim, 2013: A Bayesian approach to probabilistic streamflow forecasts. *J. Hydroinform.*, **15**, 381–391, doi:10.2166/hydro.2012.080.
- Webster, P. J., and S. Yang, 1992: Monsoon and ENSO: Selectively interactive systems. *Quart. J. Roy. Meteor. Soc.*, **118**, 877–926, doi:10.1002/qj.49711850705.
- White, W. B., 1995: Design of a global observing system for gyrescale upper ocean temperature variability. *Prog. Oceanogr.*, **36**, 169–217, doi:10.1016/0079-6611(95)00017-8.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.
- Wittenberg, A. T., A. Rosati, T. L. Delworth, G. A. Vecchi, and F. Zeng, 2014: ENSO modulation: Is it decadal predictable? *J. Climate*, **27**, 2667–2681, doi:10.1175/JCLI-D-13-00577.1.
- Wyrski, K., 1973: Teleconnections in the equatorial Pacific Ocean. *Science*, **180**, 66–68, doi:10.1126/science.180.4081.66.
- , 1975: El Niño—The dynamic response of the equatorial Pacific Ocean to atmospheric forcing. *J. Phys. Oceanogr.*, **5**, 572–584, doi:10.1175/1520-0485(1975)005<0572:ENTDRO>2.0.CO;2.
- Xie, S.-P., 2004: Satellite observations of cool ocean–atmosphere interaction. *Bull. Amer. Meteor. Soc.*, **85**, 195–208, doi:10.1175/BAMS-85-2-195.
- Xue, Y., M. Chen, A. Kumar, Z.-Z. Hu, and W. Wang, 2013: Prediction skill and bias of tropical Pacific sea surface temperatures in the NCEP Climate Forecast System version 2. *J. Climate*, **26**, 5358–5378, doi:10.1175/JCLI-D-12-00600.1.
- Zhang, S., M. J. Harrison, A. Rosati, and A. Wittenberg, 2007: System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Mon. Wea. Rev.*, **135**, 3541–3564, doi:10.1175/MWR3466.1.
- Zhang, W., Y. Leung, F.-H. Graf, and M. Herzog, 2012: Different El Niño types and tropical cyclone landfall in East Asia. *J. Climate*, **25**, 6510–6523, doi:10.1175/JCLI-D-11-00488.1.
- , F.-F. Jin, J.-X. Zhao, L. Qi, and H.-L. Ren, 2013: The possible influence of a nonconventional El Niño on the severe autumn drought of 2009 in southwest China. *J. Climate*, **26**, 8392–8405, doi:10.1175/JCLI-D-12-00851.1.
- , Y. Leung, and K. Fraedrich, 2015: Different El Niño types and intense typhoons in the western North Pacific. *Climate Dyn.*, **44**, 2965–2977, doi:10.1007/s00382-014-2446-4.
- , and Coauthors, 2016a: Improved simulation of tropical cyclone responses to ENSO in the western North Pacific in the high-resolution GFDL HiFLOR coupled climate model. *J. Climate*, **29**, 1391–1415, doi:10.1175/JCLI-D-15-0475.1.
- , and Coauthors, 2016b: Simulated connections between ENSO and tropical cyclones near Guam in a high-resolution GFDL coupled climate model: Implications for seasonal forecasting. *J. Climate*, **29**, 8231–8248, doi:10.1175/JCLI-D-16-0126.1.
- Zhao, M., H. H. Hendon, O. Alves, G. Liu, and G. Wang, 2016: Weakened eastern Pacific El Niño predictability in the early twenty-first century. *J. Climate*, **29**, 6805–6822, doi:10.1175/JCLI-D-15-0876.1.
- Zheng, F., and J. Zhu, 2016: Improved ensemble-mean forecasting of ENSO events by a zero-mean stochastic error model of an intermediate coupled model. *Climate Dyn.*, **47**, 3901–3915, doi:10.1007/s00382-016-3048-0.
- Zhu, J., B. Huang, Z.-Z. Hu, J. L. Kinter III, and L. Marx, 2013: Predicting US summer precipitation using NCEP Climate Forecast System version 2 initialized by multiple ocean analyses. *Climate Dyn.*, **41**, 1941–1954, doi:10.1007/s00382-013-1785-x.
- , A. Kumar, B. Huang, M. A. Balmaseda, Z.-Z. Hu, L. Marx, and J. L. Kinter III, 2016: The role of off-equatorial surface temperature anomalies in the 2014 El Niño prediction. *Sci. Rep.*, **6**, 19677, doi:10.1038/srep19677.